

3rd Grand Canvas : AI品質の未来を共に描く

LT

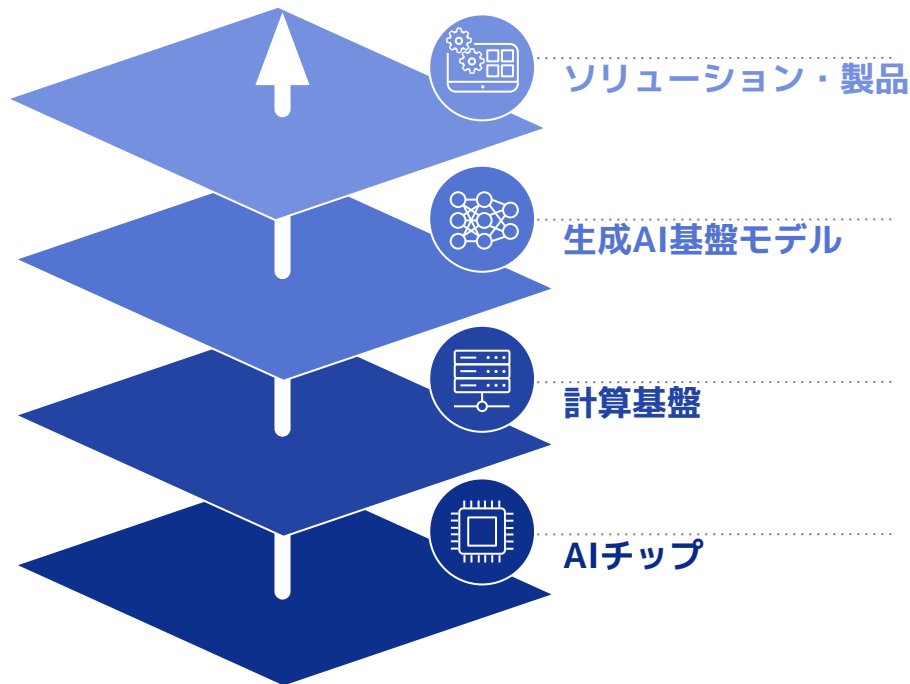
株式会社 Preferred Networks

プロダクトサービス担当 VP 兼 ビジネス戦略担当 VP 経営企画

福田 昌昭

PFNの事業: AI技術のバリューチェーンを垂直統合

PFNは、チップ、計算基盤、生成AI基盤モデル、ソリューション・製品まで、AI技術のバリューチェーンを垂直統合し、ソフトウェアとハードウェアを高度に融合することで、競争力の高い技術の開発および産業応用を進めています。



様々な産業・消費者向けのソリューション・製品



PLaMo

PLaMo Prime (今秋提供予定のLLM)
PLaMo Lite (エッジ向けSLM)

PFP

物質のエネルギー計算モデル



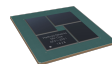
GPUクラスタ



MN-3
(MN-Core™
クラスタ)

PFCP

MN-Core™ 2を
計算資源とした
クラウドサービス



MN-Core™



MN-Core™ 2



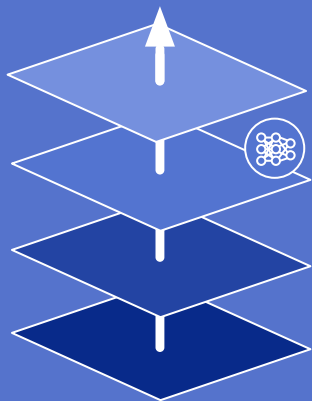
生成AI (推論) 向け
MN-Core L1000
(2026年提供予定)



MN-Core
第三世代



生成AI・ 基盤モデル



PLaMo™

PFNグループの生成AI基盤モデル

PLaMoは、PFNのグループ会社Preferred Elements (PFE)がフルスクラッチで開発する純国産の生成AI基盤モデルです。高品質な学習データを用い、既存の大規模言語モデルをベースに用いずに全く新規に開発。主要な日本語ベンチマークで世界最高レベルの精度を示しています。



- 商用版PLaMo Primeを2024年中にリリース予定
- **エッジ向け小規模言語モデル PLaMo Lite を提供中**
- 金融・医療などの特定タスクに強いモデルも順次リリース予定

世界最高クラスの日本語性能

- 主要な日本語ベンチマーク*でGPT-4を超える精度を記録
- 現在も精度向上中
- 日英の翻訳性能にも優れる

*日本語LLMの性能を測るため標準的に使われるベンチマーク「Jaster」を使用

純国産フルスクラッチモデル

- 独自のアーキテクチャと学習データでゼロから事前学習・指示学習を実施
- 他社モデルベースではないため社外ライセンスの縛りや開発上で不明瞭な点がない

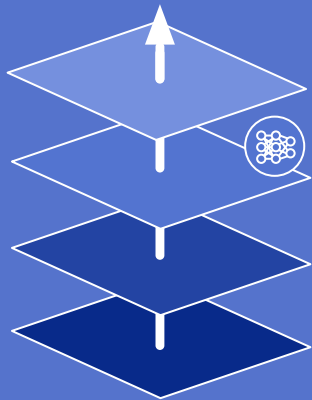
APIで簡単に導入可能

- API経由でクラウドを介して提供、簡単なコードの書き換えで実装可能
- モデル自体をお客様のクラウドなオンプレミス環境に実装することも可能

*PLaMo-100Bは、子会社のPreferred ElementsがNEDO（国立研究開発法人新エネルギー・産業技術総合開発機構）の助成事業「ポスト5G情報通信システム基盤強化研究開発事業」（JPNP20017）に採択され、日本の基盤モデル開発力向上を目指すGENIACプロジェクトで開発しています。



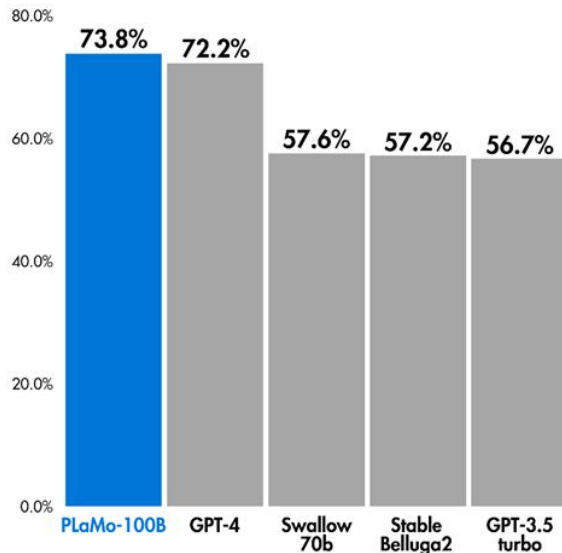
生成AI・ 基盤モデル



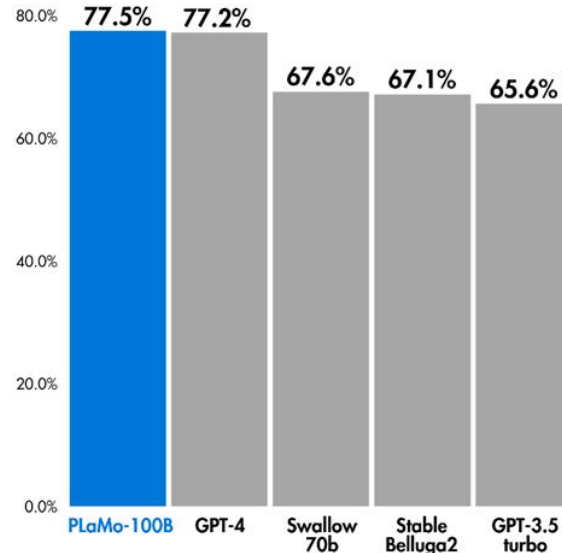
PLaMo™ β版

日本語性能世界トップレベルの大規模言語モデル

Jaster 0shot



Jaster 4shot

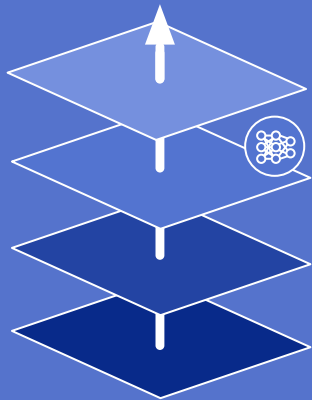


※Jaster：日本語LLMの性能を測る標準ベンチマーク

※これらの数値はGENIAC内での評価値



生成AI・ 基盤モデル

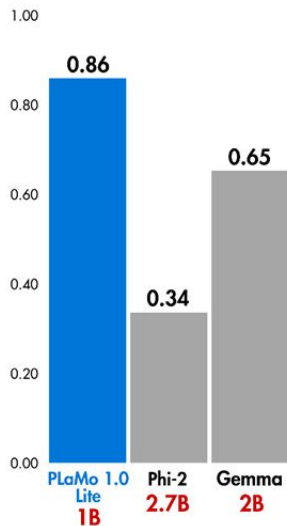


PLaMo™ Lite

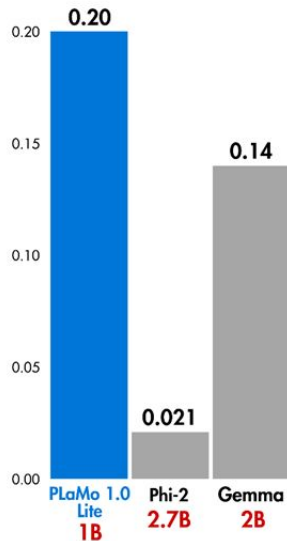
エッジデバイス向けの小規模言語モデル

- エッジデバイスで動かせる10億パラメータの小規模言語モデルの提供開始
- PLaMo-100Bの開発で培った技術に応用
- 同程度のサイズのモデルと比べて世界最高レベルの日本語性能を達成

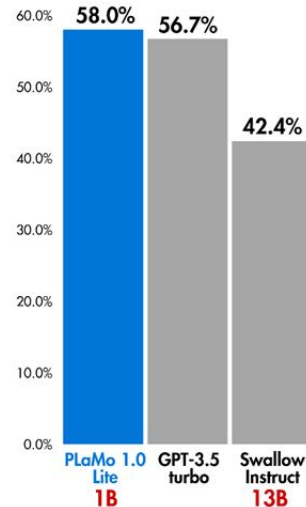
JSQuAD F1
(日本語質問応答)



WMT20 4shot
(日英翻訳 En-Ja)



Jaster 0shot



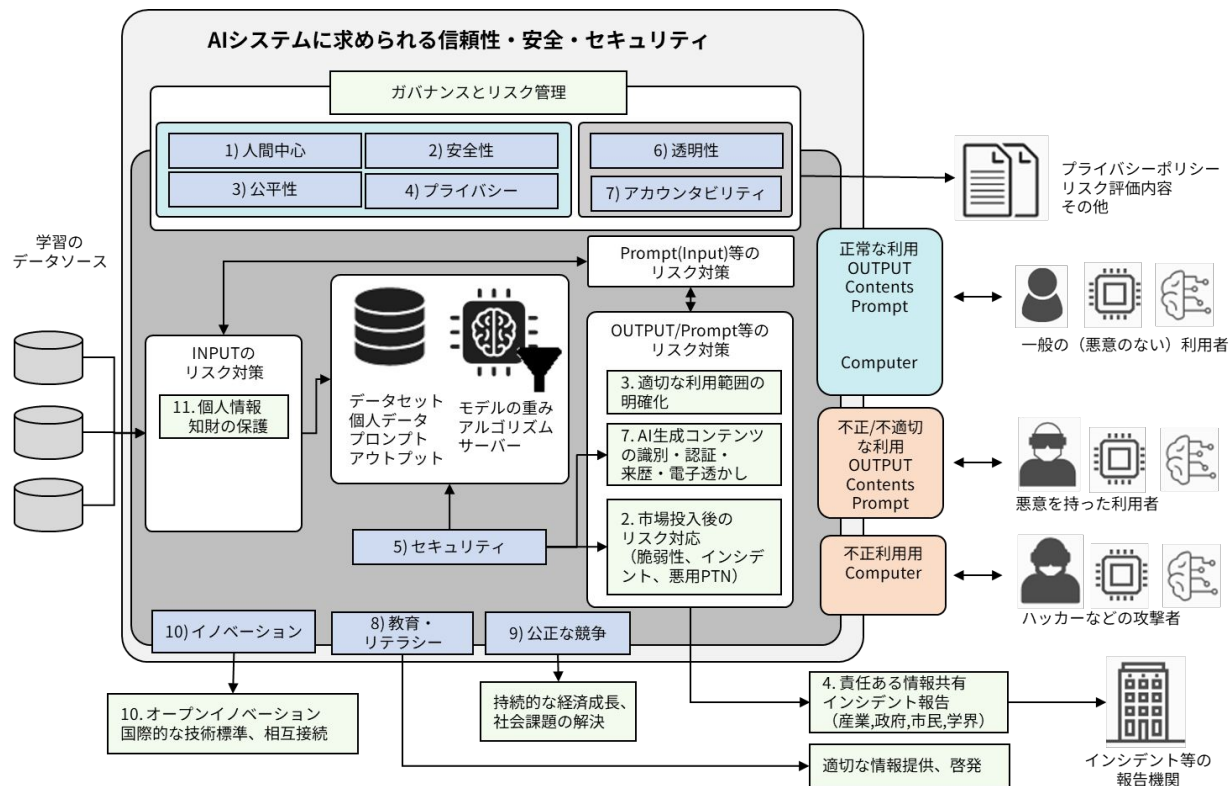
METIガイドライン、広島宣言から

基本理念

- (a) 人間の尊厳が尊重される社会 (Dignity)
- (b) 多様な背景を持つ人々が多様な幸せを追求できる社会 (Diversity & Inclusion)
- (c) 持続可能な社会 (Sustainability)

主要な想定リスク

- 兵器の開発
- 設計取得
- 参入障壁の減損
- 化学・生物・放射能・核リスク
- 攻撃的なサイバー能力
- 健康・安全
- 物理的な（制御）システムへの影響
- 自律的な複製、成長、発展
- 偏見や差別
- 法的枠組みへの違反
- 偽情報の助長
- プライバシーの侵害
- 人権に対する脅威
- 重大事故につながる連鎖



PFN AIポリシー <https://www.preferred.jp/ja/company/aipolicy/>

PFNは現実世界を計算可能にすることで、これまで解決が困難であった課題解決を目指します。人々のより豊かな生活の実現のために、生産性の向上だけでなく、人間の能力や創造性を拡大するとともに、一人ひとりの困難の克服をサポートし、これまでの人々の英知を受け継いだ上でこれを拡張し、尊厳を持って豊かな生活の実現に寄与することを目指します。

AI技術自体を支える技術領域の革新に取り組み、AIを利活用するだけでなくこれらのAIを支えるための省電力な計算力の構築により持続可能な社会の実現を目指します。そして、これらの取り組みは、企業や個人の選択枝を確保し、公正な競争を実現することにつながるものと考えています。

一方で、新技術は、強力で汎用性が高ければ様々な場面で使われるようになります。そのような技術の責任ある開発者となるため、私たちは以下の方針を守ります。

1. 透明性(正しく伝える)

私たちは、その技術自体が何をするか、何ができるか、何ができないかを可能な限り明確にします。それがどのような仕組みで動くかを、論文その他の出版、ブログ、ドキュメント等を通して、様々なレベルのステークホルダーがその技術の利用について適切に判断できるよう理解できるよう、丁寧に説明します。

2. リスク(適切に怖がる)

私たちは、その技術を使うことによって、社会の様々な場面で起きうるリスクを、私たちの想像力が及ぶ限り考慮します。リスクには、定量化できないもの、想定しにくいものがあり、また、ある時点での社会通念ではリスクではなかったものが、その後リスクと考えられるようになることもあります。新技術は常に便益とリスクを伴います。私たちは、社会との対話を通して、そのようなリスクに対応していく努力を惜しみません。

3. インテグリティ(見たくないものを、見る)

新技術の開発や利用にあたっては、時として私たちに都合の悪い発見があるかもしれません。私たちは、そのようなものから目を背けることなく、誠実に対応します。

PFNは、責任ある開発者であるため、個人の生命、自由、尊厳、財産などを脅かすことのない開発であること、推論に偏りや誤りが生じる可能性を認識し、公平性を保つように努め、多様性のある社会の実現を目指します。

データソースや利用者のプライバシーに配慮するとともに、サイバーセキュリティを確保し、AI領域特有のセキュリティに取り組めます。

ここに述べた取り組みを進めるにあたって社内で価値と技術を共有し、技術、倫理、プライバシー、セキュリティ、コンプライアンスなどに対する教育と啓発に取り組めます。会社バリューの一つであるLearn or Dieの精神の元に技術の可能性やリスクについて学び続け、これらを社会に広げていきます。